Theoretical Prediction of the Hydrogen-Bond Basicity pK_{HB}

Olivier Lamarche and James A. Platts*[a]

Abstract: Ab initio and DFT calculations on around 65 hydrogen bond or Lewis bases and their complexes with hydrogen fluoride have been performed, and a range of calculated properties from both free bases and complexes correlated with pK_{HB}, an experimental scale of hydrogen-bond basicity. For the entire range of bases, we found that the hydrogen-bond binding Gibbs free energy computed at the B3LYP/6-31+G(d,p) level of theory linearly correlated with pK_{HB}. Further improvements in the correlation and prediction of pK_{HB} were possible with a non-linear fit by considering the hydrogen bonding Gibbs free energy of another possible stereoisomeric 1:1 complex and/or that of a linear 2:1 complex, which included a second hydrogen fluoride.

Keywords: basicity · density functional calculations · Gibbs free energy · hydrogen bonds · statistics

Introduction

Hydrogen bonds are of undoubted importance throughout chemistry,[1] biology,[2] and materials science.[3] Their utility comes from their strong directional preference, which sets them apart from other intermolecular interactions, and their intermediate strength, placing them between covalent and pure van der Waals interactions.[4] Solvation is a case in point,[5] water being the most obvious example of a solvent where hydrogen bonding to solute molecules is important. Phenomena such as passive diffusion (as opposed to active or carrier-mediated transport) through biological liquids or membranes,[6] partition between phases[7] and adsorption[8] are similarly dependent to some extent on hydrogen bonding.

Reflecting this importance, several experimental scales of hydrogen-bond acidity and basicity have been proposed. Taft's pK_{HB} scale was the first example of such an attempt.[9] The pK_{HB} value is related to the Gibbs energy ΔG^0 of formation of the hydrogen-bonded complexes in CCl_4 at 298 K [Eqs.(1) – (3)].

$$4\text{-}FC_6H_4OH + B \rightleftharpoons 4\text{-}FC_6H_4OH\cdots B \quad (1)$$

$$K_f = [\text{Complex}]/[\text{Base}][4\text{-}FC_6H_4OH] \quad (\text{in dm}^3\,\text{mol}^{-1}) \quad (2)$$

$$pK_{HB} = \log K_f = \Delta G^0/5.7 \quad (\Delta G^0 \text{ in kJ mol}^{-1}) \quad (3)$$

The equilibrium constant was calculated from the equilibrium concentration of acid,[10] which was obtained from IR absorbance of the free OH band at 3614 cm^{-1} (Beer – Lambert law).

The development of this basicity scale came up against various experimental difficulties. Bases not commercially available, gaseous, insufficiently soluble in CCl_4 and/or self-associated cannot be easily studied. Moreover, this can be a very costly and time-consuming process. We therefore sought alternative methods to calculate pK_{HB} values using the techniques of quantum chemistry, with the ultimate aim of calculating pK_{HB} for any molecule purely from structure.

Le Questel and co-workers reported a semiempirical approach for estimating pK_{HB}.[11] They selected three descriptors, the enthalpy of formation of the bimolecular hydrogen-bond reaction, the hydrogen-bond length and the most negative value of the electrostatic potential. Using descriptors computed at AM1 level for hydrogen-bond formation between 4-fluorophenol and the base, the best correlation that they obtained was with the enthalpy of formation ($n = 22$, $R^2 = 0.951$, rms = 0.17). However, they restricted the diversity of compounds to nitrile types only. Using the molecular electrostatic potential, generalisation of the model had been attempted,[12] but the fit was rather poor ($n = 33$, $R^2 = 0.81$, rms = 0.39) as the descriptor was computed for a variety of acceptor heteroatoms and therefore became family-dependent.[13] It is worth noting that a good multivariate correlation for the OH frequency shifts of methanol had been found[12] ($R^2 = 0.966$) and the predicted values could then be correlated to the pK_{HB}, by using the empirical relationships developed by Laurence and co-workers ($pK_{HB} \propto \Delta\tilde{\nu}_{OH}$).[14] However, this model is experimentally dependent and the predicted values depend on a sequence of correlations, which could lead to cumulative errors. Our aim was to find a general model, which

[a] Dr. J. A. Platts, O. Lamarche
Department of Chemistry, Cardiff University
P.O. Box 912, Cardiff CF10 3TB (UK)
Fax: (+44) 2920 874030
E-mail: platts@cf.ac.uk

  0947-6539/02/0802-0457 $ 17.50+.50/0

would give a quick and easy way to predict a large range of pK_{HB} values for a wide diversity of compounds, purely from theoretical calculations.

A more general β_2^H scale[15] was recently developed by Abraham and co-workers. Similarly to pK_{HB}, this was an equilibrium constant for complexation of bases with a reference acid measured in a non-hydrogen bonding solvent. However, the scale is defined such that zero is no hydrogen-bonding ability and one is a strong base (such as hexamethylphosphorous triamide).

Thus, β_2^H is a free energy property, but could be expressed in conventional units. The β_2^H scale (referring to 1:1 complexation) was then generalised to be relevant to the solvation situation in which a solute was surrounded by solvent molecules and hence underwent multiple hydrogen-bond formation. Using a number of multiple linear regression equations, the β_2^H values were back-calculated, resulting in the "effective" or "summation" $\Sigma\beta_2^H$ scale.[16]

A previous study[17] showed that several density functional theory (DFT) calculated properties for complexes with hydrogen fluoride correlated with $\Sigma\beta_2^H$ with reasonable accuracy. These properties include: the hydrogen-bond binding energy, changes in electron density at critical points and the lengthening and weakening of the H–F bond on hydrogen-bond formation. In the present study, we sought similar models for the hydrogen-bond basicity scale pK_{HB}.

Reliable calculation of descriptors by DFT methods with large basis sets seemed unlikely to be realised for extended sets of relatively large molecules complexed with 4-fluorophenol. As an alternative, we used complexes with hydrogen fluoride, the smallest possible hydrogen-bond acid, which dramatically reduced the computational time and the chance of secondary interactions complicating matters. This approach was already used with success in the calculation of descriptors for $\Sigma\beta_2^H$. Furthermore, Abraham and co-workers[15] have reported linear equations of $\log K$ values for the complexation of a series of bases (i) [see Eq. (4)] against a number of reference acids in dilute solution in tetrachloromethane [Eq. (4), L_A and D_A characterise the acids].

$$\log K^{(i)} = L_A \cdot pK_{HB}{}^{(i)} + D_A \quad (4)$$

The error in experimentally determined pK_{HB} values has been estimated at around 0.05 units,[18] as the K_f value was usually estimated to be accurate to within 5–10%. An experimental standard error of 0.10 units should be considered for our purposes as we have a smaller amount of data.

Computational Methods

A set of 40 hydrogen-bond bases was selected from Laurence's experimental database, values being obtained by the experimental methods outlined above. These molecules were chosen to cover both the numerical spread and the chemical diversity of molecules with known pK_{HB} values of simple bases, from very weak S and π-acceptor atoms, through intermediate O acceptors, to very strong amines and other N bases. The molecules are tabulated in Table 1, along with a sequential number and their experimental pK_{HB} value. No restriction on the number of basic sites within a molecule was made. Another set of 25 compounds was selected to test the models so developed: molecules were selected to cover known or suspected weaknesses of the model. The molecules are tabulated in Tables 5 and 7 (see below), along with a sequential number and their experimental and observed pK_{HB} values.

Table 1. Values of the hydrogen-bond basicity parameter for the training set.

Number	Name	pK_{HB}
alcohol[a]		
1	4-fluorophenol	– 0.13
2	phenol	– 0.07
3	methanol	0.82
4	ethanol	0.96
alkyl halide[b]		
5	bromoethane	– 0.40
6	2-chloropropane	– 0.30
amine[c]		
7	trimethylamine	1.93
8	ethylamine	2.17
9	methylamine	2.20
10	dimethylamine	2.23
11	*tert*-butylamine	2.26
12	pyrrolidine	2.56
aniline[d]		
13	*N*-methylaniline	0.40
14	aniline	0.56
carboxylic amide[e]		
15	methyl formamide	1.96
16	methyl acetamide	2.30
17	dimethyl acetamide	2.44
carboxylic ester[f]		
18	ethyl formate	0.66
19	methyl benzoate	0.89
20	ethyl acetate	1.07
ether[g]		
21	tetrahydropyran	1.23
ketone[h]		
22	acetone	1.18
nitrile[i]		
23	chloroacetonitrile	0.39
24	acrylonitrile	0.70
25	methyl thiocyanate	0.73
26	acetonitrile	0.91
π-base[j]		
27	hept-1-ene	– 0.67
28	benzene	– 0.49
29	hex-1-yne	– 0.22
30	hex-3-yne	– 0.10
31	*N*-methylpyrrole	– 0.07
pyridine[k]		
32	3-chloropyridine	1.31
33	4-chloropyridine	1.54
34	pyridine	1.86
35	4-methylpyridine	2.07
36	4-aminopyridine	2.56
sulfide[l]		
37	thioanisole	– 0.10
38	dimethyl sulfide	0.12
sulfoxide[l]		
39	dimethyl sulfoxide	2.53
urea[e]		
40	1,1,3,3-tetramethylurea	2.44

[a] C. Laurence, *J. Phys. Chem.* **1989**, *93*, 3799. [b] C. Ouvrard, *J. Chem. Soc. Perkin Trans. 2* **1999**, *7*, 1357. [c] J. Graton, *J. Chem. Soc. Perkin Trans. 2* **1999**, *5*, 997. [d] C. Laurence, unpublished results. [e] J. Y. Le Questel, *J. Chem. Soc. Perkin Trans. 2* **1992**, *12*, 2091. [f] F. Besseau, *J. Chem. Soc. Perkin Trans. 2* **1994**, *3*, 485. [g] M. Berthelot, *Eur. J. Org. Chem.* **1998**, 925. [h] F. Besseau, *J. Chem. Soc. Perkin Trans. 2* **1998**, *1*, 101. [i] J. Y. Le Questel, *J. Chem. Soc. Perkin Trans. 2* **1997**, *12*, 2711. [j] F. Besseau, *Bull. Soc. Chim. Fr.* **1996**, *133*, 381. [k] F. Besseau, *J. Chem. Soc. Perkin Trans. 2* **1998**, *2*, 283. [l] R. W. Taft, *J. Am. Chem. Soc.* **1969**, *91*, 4801.

0947-6539/02/0802-0458 $ 17.50+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 2

All ab initio and DFT calculations were performed using GAUSSIAN98[19] running on a Compaq XP1000 workstation. Initially, the geometries of the bases were optimised at the HF/6-31G(d,p) level,[20] and the resulting structures were confirmed as minima through harmonic frequency calculations. Where it was deemed necessary, the conformational space of the molecules was explored at the same level to ensure the final optimised structure corresponded to the global minimum. Taking these HF/6-31G(d,p) geometries as starting points, hydrogen-bonded complexes with hydrogen fluoride were generated by placing the H–F molecule approximately 1.75 Å from the acceptor (denoted B) nucleus in the expected acceptor region (i.e., the lone-pair or π system). The geometries of these complexes were again optimised at the HF/6-31G(d,p) level.

A number of properties outlined below were then computed as possible descriptors of pK_{HB} by performing single-point calculations at the B3LYP/6-311G(d,p) level[21–22] using the HF/6-31G(d,p) geometries. The geometries of these molecules and complexes were then reoptimised at the B3LYP/6-31+G(d,p)[23] level and the properties were recomputed at the same level in order to obtain more accurate energies and properties. Single point PCM-UAHF[24] calculations were performed at the B3LYP/6-31+G(d,p) level in order to study the solvent effect. Subsequently, all the other 1:1 and 2:1 complexes involved in further studies were also optimised at the B3LYP/6-31+G(d,p) level.

We have tested two descriptors that were successful in modelling $\Sigma\beta_2^H$: the hydrogen-bond binding energy and the lengthening and weakening of the H–F bond on hydrogen-bond formation. Four extra descriptors were added for the B3LYP geometries, namely the enthalpy and Gibbs free energy of hydrogen bond formation at 298 K, ΔH^0 and ΔG^0 (thermal energy and entropy terms were calculated on the assumption of harmonic behaviour around all normal modes). The third was the hydrogen-bond length $r(B\cdots H)$, successfully used by Le Questel,[11] and the last was the IR frequency shift of the HF band on complexation, defined as $\Delta\tilde{\nu} = 4068 - \tilde{\nu}(HF_{complex})$. It is known that the correlation between pK_{HB} and the frequency shifts of the OH band of 4-fluorophenol is usually excellent within families.[14]

The computed properties were used as independent variables in a standard least-squares regression. All simple linear regression analysis employed the JMP discovery software, published by SAS software.[25] The significance or otherwise of calculated properties in the correlation of pK_{HB} was determined using the standard t-test and the predictive ability of any model tested using the "leave-one-out" cross-validation method. Statistical measures of correlations used were: R^2, the overall correlation coefficient equal to the percentage of variance accounted for in the model; R_{CV}^2, the cross-validated or leave-one-out R^2 value, a measure of the predictive ability of the model; rms, the root-mean-square error; and Fischer's *F*-statistic, a measure of the significance of the model given the number of variables employed.

Results and Discussion

The ab initio and DFT approach to 1:1 complexes

Models of pK_{HB} from the most stable 1:1 complex: The first approach was to use the method developed for its analogue, $\Sigma\beta_2^H$. In effect, the hydrogen-bonding descriptor, β_2^H, was just a more convenient scale[15–16] of pK_{HB} and could be easily deduced from it, using Equation (5).

$$\beta_2^H = (pK_{HB} + 1.1)/4.636 \quad (5)$$

Although the two scales are related, there is no direct way to calculate pK_{HB} from $\Sigma\beta_2^H$ as the latter is deduced by using a number of multiple linear regression equations and by "back-calculating" its value from the solute hydrogen-bond parameter, β_2^H. However, these two parameters are usually similar and it was therefore expected that the same descriptors to model them could be used.[16]

The first part of Table 2 reports statistics for fits of two calculated properties against experimental pK_{HB} values. Correlation of pK_{HB} against the hydrogen-bond binding energy (ΔE_{el}, in kJ mol^{-1}) and the lengthening and weakening of the H–F bond on hydrogen-bond formation (Δr(H–F) in au) resulted in surprisingly poor statistical data, considering that these properties were the most successful descriptors of $\Sigma\beta_2^H$.[17] The different nature of these scales ($\Sigma\beta_2^H$ scale was back calculated) had increased their scattering ($n = 40$, $R^2 = 0.824$). The quality of the statistics shown in the first part of Table 2 could suggest that hydrogen-bond complex formation was not an iso-entropic reaction between compounds or between families.

Table 2. Linear correlations with HF complex properties.

Model[a]	*n*	R^2	R_{CV}^2	rms	*F*-stat
HF geometry/B3LYP property					
Δr(H–F)	40	0.834	0.819	0.423	191.6
ΔE_{el}	40	0.858	0.842	0.392	229.9
B3LYP geometry/B3LYP property					
$r(B\cdots H)$	40	0.658	0.627	0.609	73.00
$\Delta\tilde{\nu}$	40	0.744	0.721	0.527	110.4
ΔE_{el}	40	0.906	0.897	0.319	366.5
ΔE_{PCM}	40	0.883	0.872	0.356	285.7
ΔH^0	40	0.911	0.902	0.311	386.7
ΔE_{el}+ΔZPE	40	0.922	0.915	0.290	450.7
$\Delta E_{el}+\Delta E_{rtv}$	40	0.894	0.883	0.339	320.6
$\Delta E_{el} - T\Delta S$	40	0.918	0.911	0.298	425.9
$\Delta G^0 - \Delta E_{rtv}$	40	0.926	0.919	0.283	476.4
ΔG^0	40	0.928	0.922	0.279	491.9

[a] See text for definition.

In an attempt to improve the correlation, the geometry of the molecules and their corresponding complexes were reoptimised at a higher and more flexible level of theory: the B3LYP/6-31+G(d,p). The second part of Table 2 reports statistics for fits of calculated properties against experimental pK_{HB} values: the hydrogen-bond binding energy, enthalpy and Gibbs free energy (ΔE_{el}, ΔH^0 and ΔG^0, in kJ mol^{-1}), the hydrogen bond length ($r(B\cdots H)$ in au) and the IR frequency shift of the HF band (in cm^{-1}). The correlation between pK_{HB} and the hydrogen bond length is very poor. By plotting the data, we can clearly see that their relationship is not linear and the correlation can be improved to $R^2 = 0.853$ by using an equation of the form: $a + b/(R + c)$. The correlation between pK_{HB} and $\Delta\tilde{\nu}$ was not very good and seemed to be family-dependent.

The best model came from the Gibbs free energy, which correlated well with pK_{HB}, according to Equation (6):

$$pK_{HB} = 0.481\,(0.051) + 0.080\,(0.004)\cdot\Delta G^0 \quad (6)$$

with $n = 40$, $R^2 = 0.928$, $R_{CV}^2 = 0.922$, rms = 0.279, $F = 491.9$. Already, the improvement of the correlation between pK_{HB} and ΔE_{el} was dramatic between the two methods of calculation, showing the importance of the level of theory for accurate and error-consistent calculation of energy. As the pK_{HB} was measured in CCl_4, calculations were also performed in the presence of this solvent, using the polarized continuum (overlapping spheres) model (PCM). The PCM hydrogen-

bond binding energies obtained, ΔE_{PCM}, were close to the gas-phase ones and gave similar correlation, but with slightly more scattering. Gibbs free energies were not recomputed using PCM as the computational time was too great and as we had no guarantee of improvement.

A careful study of the different components between ΔE_{el} and ΔG^0 (hydrogen-bond binding zero point energy: ΔZPE, hydrogen-bond binding entropy: ΔS and hydrogen-bond binding energy of vibration, translation and rotation: ΔE_{rtv}) showed that the main improvement comes from the terms: ΔS and ΔZPE. In effect, according to Table 2, the fit of ΔE_{el} was only improved by the introduction of ΔZPE or $T\Delta S$ (with $T = 298$ K). The introduction of ΔE_{rtv} had a negative impact on the model of ΔE_{el}, which explains why ΔE_{el} and ΔH^0 gave equivalent models. However, the omission of the thermal term ΔE_{rtv} into ΔG^0 did not improve the statistics of the latter.

Calculations of ΔS values showed that the formation of hydrogen bond was iso-entropic between compounds of the same family (variation is < 7 J mol^{-1} in our data set, if aromatic and alkane derivatives were considered separately) but not between families (differences of up to 60 J mol^{-1}, equivalent to $\Delta pK_{HB} = 1.4$). The fact, that the hydrogen-bond Gibbs free energy was the best descriptor of pK_{HB}, is not really surprising as, by definition, the hydrogen-bond basicity is related to the Gibbs energy of formation of the hydrogen-bonded complexes [Eq. (3)].[9] Figure 1 is a plot of the observed versus calculated pK_{HB} values from Equation (6).

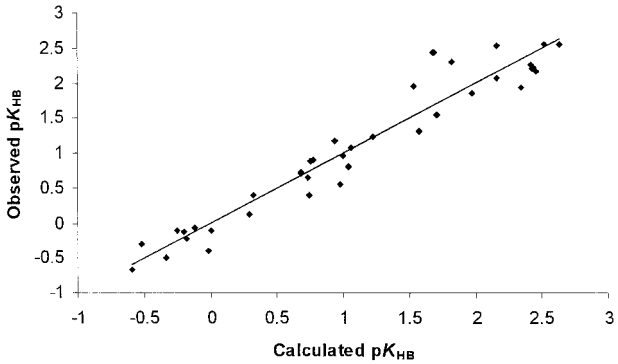

Figure 1. Observed versus calculated pK_{HB} from Equation (6).

Contributions from 1:1 complex on second hydrogen-bond site: Tables 1 and 5 contain the apparent or global pK_{HB} value, meaning the value obtained from the equilibrium concentration of acid. However, the K_f is the global formation constant which is the sum of the formation constants for each possible complex. In effect, experimental correction on the formation constant could be applied to take into account a possible second acceptor atom. For example, with 4-fluorophenol as acid, the apparent value of aniline is 0.56 and the effect of the π-base can be subtracted (evaluated by the IR spectra and $pK_{HB} = f(\Delta\tilde{\nu})$ relationship) to obtain the true amine pK_{HB} value, 0.49.[26]

To test the hypothesis that the best correlation is between the apparent value and the most stable 1:1 complex, we performed calculations of hydrogen bonding on the other possible basic sites of molecules, as displayed in Table 3. The first Gibbs free energy was that of the most stable 1:1 complex, and the second that of the other possible site. From these two Gibbs free energy and Equation (6), we could deduce an apparent pK_{HB} value, app. pK_{HB}, by summing their respective formation constant (details of these calculations are in the Section on DFT calculations, see below). It is clear that the difference between the two pK_{HB} values was insignificant, that is the formation constant of the second site could be omitted. Therefore, we considered only the formation constant of the complex from the most stable site. It seemed that the error on our prediction of pK_{HB} was much greater than any correction that may have been necessary to the formation constant.

Table 3. Impact of other hydrogen-bond sites on pK_{HB} prediction.

Number	Name	ΔG_1^{0}[a]	ΔG_2^{0}[b]	Type[c]	pK_{HB}[d]	app. pK_{HB}[e]
compounds from the training set						
36	4-aminopyridine	25.41	−2.11	Nsp^3	2.51	2.52
14	aniline	6.24	−12.98	π	0.98	0.99
23	chloroacetonitrile	−2.01	−19.19	Cl	0.32	0.34
17	dimethyl acetamide	14.87	−20.49	Nsp^2	1.67	1.67
39	dimethyl sulfoxide	20.99	−17.33	S	2.16	2.16
25	methyl thiocyanate	2.51	−19.89	S	0.68	0.69
15	methyl formamide	13.17	−21.51	Nsp^2	1.53	1.54
34	pyridine	18.67	−20.22	π	1.97	1.97
40	1,1,3,3-tetramethylurea[f]	15.08	−4.30	Nsp^2	1.69	1.71
compounds from the prediction set						
44	dimethylcarbamoyl chloride	1.23	−30.51	Nsp^2	0.58	0.58
51	2-chloroethylethyl ether	0.69	−15.05	Cl	0.54	0.56
59	dimethyl cyanamide	11.22	−13.79	Nsp^3	1.38	1.38

[a] Hydrogen-bond Gibbs free energy of the most stable complex, in kJ mol^{-1}. [b] Hydrogen-bond Gibbs free energy of the second hydrogen-bond site, in kJ mol^{-1}. [c] Type of base of the second hydrogen-bond site. [d] From Equation (6) and ΔG_1^0. [e] pK_{HB} value taking into account the second hydrogen-bond site, the apparent pK_{HB} (cf. discussion). [f] The second formation constant has been used twice as there are two nitrogen atoms in this molecule.

However, in the case of identical sites (such as symmetrical compounds), the values of the formation constant were identical. Therefore, since K_f values were added, the pK_{HB} value had to be statistically modified by $-\log 2$ (-0.301). In our training set, the value of *N*-methylpyrrole had been modified by $-\log 2$ as H–F was bound to one double bond (see Figure 2, hydrogen bonding on the nitrogen atom was a lot less stable as it destroyed the planarity of the molecule). Other compounds with similar corrections have been introduced later in the prediction test (such as pyrimidine and biacetyl).

Position of hydrogen bonding to sp^2 and sp^3 hybrid oxygen atoms: Throughout the above discussion, we were concerned only with the global minimum of the 1:1 complex. However, in

0947-6539/02/0802-0460 $ 17.50+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 2

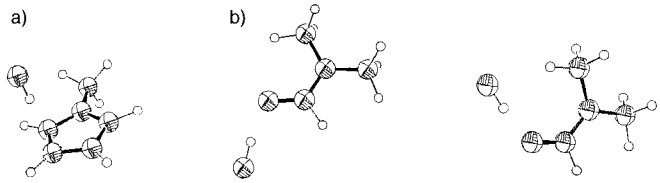

Figure 2. Examples of 1:1 complexes: a) *N*-methylpyrrole, b) dimethyl formamide (stereoisomeric complexes).

the case of sp^2 or sp^3 oxygen bases, the most stable conformation is not as obvious as that, for example, of amines. Experimentally, it was found that the $\tilde{\nu}$(OH) band of the complex of 4-fluorophenol with such bases was often broad and unsymmetrical, even for very low concentrations of acid. It could be resolved into Gauss–Lorentzian component bands[27–29] and was generally attributed to other 1:1 complexes (Figure 3). The main frequency was usually attributed to the bent n complex (**1** or **2**) and the other frequencies were attributed to linear n complex[28] (**3**) or out-of-plane π complex[29] (**4**). Calculations were performed on these different 1:1 complexes. All the complexes of types **3** and **4** were unstable, and rearranged to bent n complexes (except in the case of phenols, where the most stable configuration was of type **3**).

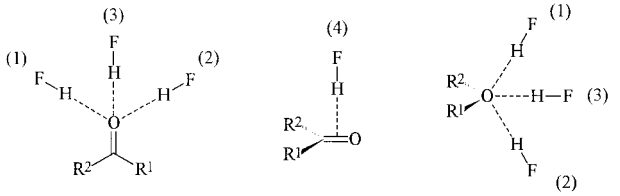

Figure 3. Different possible 1:1 complexes.

Therefore, the only stable 1:1 complexes for these types of compounds were the ones involving directly the lone pairs (except for phenols). Moreover, in the case of unsymmetrical compounds, the two lone pair Gibbs free energies were not equivalent and, contrary to the results presented in Table 3, their energies were relatively similar (identical in the case of symmetrical compound). Therefore, it was decided to study the impact of these stereoisomeric 1:1 complexes on the model.

DFT calculations of stereoisomeric 1:1 and 2:1 complexes: Even though the statistics for Equation (6) were good, the rms error (0.279) of this model was still rather high. In effect, some bases still have very high residual values (such as 1,1,3,3-tetramethylurea or dimethyl acetamide). Therefore, we decided to study the contribution of stereoisomeric 1:1 complexes (as discussed above) and 2:1 complexes to calculated values of pK_{HB}.

Discussion on the possible contribution of 2:1 complexes: Abraham et al. have studied the variation of K_f with respect to the initial acid concentration,[30] concluding the existence of a 2:1 complex (see type II in Figure 4). Optimisations of the two possible 2:1 complexes were in accordance with the experimental conclusion: the type II complex was the most stable (see Figure 5 for some examples of 2:1 complexes). At first sight, 2:1 complexes seemed irrelevant, as experimental values were determined using very low concentration of acid (compared to base) to avoid the formation of 2:1 complex and the possible dimerisation of the acid. The equilibrium constant was then taken as the mean of four values corresponding to four base concentrations.

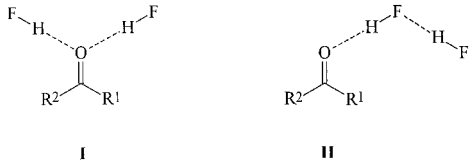

Figure 4. Two possible 2:1 complexes.

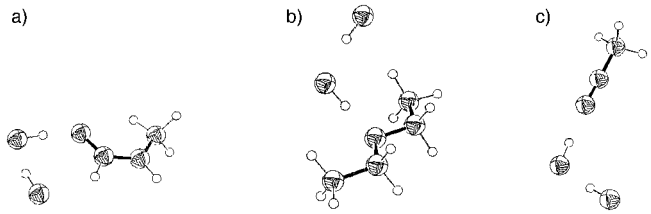

Figure 5. Examples of 2:1 complexes: a) methyl formamide, b) diethyl ether, c) acetonitrile.

However, the enthalpy of formation of the second hydrogen bond could be higher than the first, which could explain the formation of 2:1 complexes for strong bases even with low acid concentration. Preliminary calculations on acetone were encouraging: The incorporation of the linear 2:1 complex into the model increased the pK_{HB} value by 0.04 units. In effect, the calculated pK_{HB} value from Equation (6) was 0.94, but by taking into account the 2:1 complex (details of calculation in the following part), this became 0.98 (by using the lowest experimental acid concentration of 8 mM; the experimental pK_{HB} value was 1.18). So, despite the low concentration of acid used, it will be interesting to study the impact of incorporating 2:1 complexes into the scattering of the model.

Non-linear fit using the Gibbs free energy of stereoisomeric 1:1 and 2:1 complexes: In the previous calculations, the global formation constant was defined as an apparent constant, K_{app}, which was the 1:1 complex formation constant K_1 of the most stable configuration ($K_{app} = K_1$). The apparent formation constant defined by Laurence [first part of Eq. (7)] had to be modified to take into account the following contributions: K_1', the formation constant of a second, stereoisomeric 1:1 complex; K_2, the 2:1 complex formation constant for linear hydrogen bonding on the most stable 1:1 complex; and K_2', the 2:1 complex formation constant for linear hydrogen bonding on the second 1:1 complex.

In our method, we obtained an expression of the type $K_{app} = f(K_1, K_2, K_1', K_2')$ as follows. At equilibrium, let the

concentration of acid be a, that of the base be b, that of the 1:1 complexes be $[x, z]$ and that of the 2:1 complexes be $[y, w]$. Then,

$$\underset{w}{\text{FH--FH--B}} \rightleftharpoons \underset{z}{\text{FH--B}} + \underset{a}{\text{HF}} \rightleftharpoons \underset{b}{\text{B}} + \underset{a}{\text{HF}} \rightleftharpoons \underset{x}{\text{B--HF}} + \underset{a}{\text{HF}} \rightleftharpoons \underset{y}{\text{B--HF--HF}}$$

$$a_0 = a + x + z + 2y + 2w \quad \text{and} \quad b_0 = b + x + z + y + w$$

and

$$K_1 = x/ab,\ K_2 = y/ax,\ K_1' = z/ab \text{ and } K_2' = w/az$$

and K_{app} is given by the Equation (7),

$$K_{app} = \frac{a_0 - a}{a\cdot(b_0 + a - a_0)} = K_1 \cdot \left(\frac{1 + 2\cdot a\cdot K_2}{1 - a^2\cdot K_1\cdot K_2 - a^2\cdot K_1'\cdot K_2'}\right) + K_1' \cdot \left(\frac{1 + 2\cdot a\cdot K_2'}{1 - a^2\cdot K_1\cdot K_2 - a^2\cdot K_1'\cdot K_2'}\right) \quad (7)$$

The four terms of the function $K_{app} = f(K_1, K_2, K_1', K_2')$ could be calculated by computing the Gibbs free energy of their corresponding complex and using Equation (6). Unfortunately, the concentration of acid a was not known, and experimental pK_{HB} values were averaged over several a values. As this concentration was related to the equilibrium reaction with 4-fluorophenol and had no chemical sense in our model, we decided to convert the acid concentration into a regression coefficient, a_{HF}, and calculate an optimum value. The final stage was a non-linear fitting of the data of the equations below, thereby generating the optimum value of the adjustable coefficient a_{HF} and those of the formation constants [by combining Eqs. (3) and (6)].

Three cases have been studied: a) the first case considered only the most stable 1:1 complex and its stereoisomeric 1:1 complex, b) the second case considered only the most stable 1:1 complex and its corresponding 2:1 complex and c) the last case was the general one.

Case 1: K_2 and $K_2' = 0$

In order to get a systematic model, the corresponding stereoisomeric 1:1 complex had been computed for every complex with sp^2 or sp^3 hybrid oxygen or sulfur atom as the most stable acceptor site (except for the aromatic alcohols as they have linear n complexes). Therefore, from Equations (3) and (7) and by not considering any 2:1 complexes, we could deduce a non-linear fit of pK_{HB} with respect to the Gibbs free energies as displayed in Equation (8).

$$pK_{HB} = \log_{10}(K_{app}) = \log_{10}(K_1 + K_1'),$$
$$\text{with } K_1 = 10^{(\alpha\cdot\Delta G_1^{0\prime} + \beta)} \text{ and } K_1' = 10^{(\delta\cdot\Delta G_1^{0\prime} + \beta)} \quad (8)$$

The results for the two different models considered are displayed in the first part of Table 4.The correlations were obtained by computing the $\Delta G_1^{0\prime}$ values of the appropriate families in Table 1 (the corresponding formation constant term (K_1') was taken as zero if the molecules have no $\Delta G_1^{0\prime}$ value). The two models differed by their value of δ: The first model set δ equal to α, that is the calculations of the formation constants of the two 1:1 complexes were identical. This model correctly adjusted pK_{HB} values for molecules with energetically identical stereoisomers by $-\log 2$. A second model removed the constraint of $\delta = \alpha$, adding a degree of freedom to the model: both are reported in Table 4. Improvements from both models in comparison to Equation (6) were remarkable, especially in the rms error. The values of the different coefficients to calculate the formation constant are similar to Equation (6). However, for the second model, the newly introduced degree of freedom, δ, had a value larger than its equivalent, α, such that the second 1:1 complex contributed more to pK_{HB}. This model was preferred, since it resulted in rather better fit statistics (rms error 0.23 vs 0.25), even though it did not rigorously give the $-\log 2$ correction required for symmetrical complexes. Overall, we surmised that the introduction of stereoisomeric 1:1 complexes into the model improves greatly the statistics of the model, except for the Fischer's F-statistic in the case of the second model as it has an extra degree of freedom.

Case 2: K_1' and $K_2' = 0$

We have also incorporated 2:1 complexes into the model, by calculations of linear 2:1 complexes (type II in Figure 4) from

Table 4. Non-linear correlations with HF complex properties for 40 molecules.

Model[a]	Regression coefficients[b]						Statistics[c]				
	α	β	δ	a	b	n	R^2	R_{CV}^2	rms	F-stat	Max-Min
case 1: K_2 and $K_2' = 0$											
1	0.079	0.409	0.079	–	–	40	0.944	0.938	0.247	634.8	1.06
	0.003	**0.045**	**0.003**	–	–						
2	0.077	0.374	0.115	–	–	40	0.954	0.950	0.226	385.1	0.96
	0.003	**0.043**	**0.010**	–	–						
Case 2: K_1' and $K_2' = 0$											
3	0.069	0.308	–	0.041	10.89	40	0.952	0.947	0.230	369.4	0.98
	0.006	**0.052**	–	**0.008**	**3.30**						
Case 3 : General case											
4	0.057	0.225	0.057	0.037	9.58	40	0.958	0.953	0.216	420.5	0.84
	0.008	**0.066**	**0.008**	**0.011**	**4.74**						
5	0.052	0.199	0.095	0.034	10.05	40	0.968	0.965	0.191	364.6	0.81
	0.006	**0.054**	**0.014**	**0.010**	**4.33**						

[a] See text for definition. [b] The values in bold below the regression coefficients are the standard errors. [c] Max-Min is the difference between the maximum and minimum residual pK_{HB} values.

0947-6539/02/0802-0462 $ 17.50+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 2

the most stable 1:1 complexes. Only those molecules from Table 1 with positive pK_{HB} values were included in this study, as earlier calculations found that the formation constant, K_2' of negative pK_{HB} molecules was roughly zero. Therefore, from Equations (3) and (7) and by not considering any other 1:1 complexes, we can deduce a non-linear fit of pK_{HB} with respect to the Gibbs free energies as displayed in Equation (9).

$$pK_{HB} = \log_{10}\left(K_1 \cdot \left(\frac{1 + 2 \cdot a_{HF} \cdot K_2}{1 - a_{HF}^2 \cdot K_1 \cdot K_2}\right)\right), \quad \text{with } K_i = 10^{(\alpha \cdot \Delta G_i^0 + \beta)} \text{ and } a_{HF} = \frac{1}{a \cdot K_1 \cdot K_2 + b} \tag{9}$$

The operand of the logarithmic term in Equation (9) had to be positive for its $\log_{10}$ to be defined, that is a_{HF} must be less than $1/\sqrt{(K_1 \cdot K_2)}$. Simply using a_{HF} as a regression coefficient led to it approaching zero as we introduced more strong bases, cancelling the effect of the incorporation of 2:1 complexes. Thus, the adjustable coefficient a_{HF} took the form shown because it was related to concentration of acid, and hence inversely proportional to the strength of the base (K_1 and K_2). Using the Gibbs free energy of the 2:1 complexes, ΔG_2^0, we obtained the third model displayed in the second part of Table 4. The improvement on the linear model was impressive, similar to that for model 2. However, it showed some family-dependent errors, for example pyridine compounds were less well predicted, especially at low pK_{HB} value. We therefore preferred model 2 from Table 4, as no real improvements are noticeable.

Case 3: **General case**

The previous models displayed in Table 4 took into account either the stereoisomeric 1:1 complexes or the 2:1 complexes. The logical following step will be to consider a model with these complexes altogether. However, one hypothesis has been made on these two last models to avoid any excessive time and computational resources: The formation constant of the second 2:1 complex (K_2') has been taken equal to the one of the first 2:1 complex (K_2). Therefore, from Equations (3) and (7), we can deduce a non-linear fit of pK_{HB} with respect to the Gibbs free energies as displayed in Equation (10).

$$pK_{HB} = \log_{10}\left((K_1 + K_1') \cdot \left(\frac{1 + 2 \cdot a_{HF} \cdot K_2}{1 - a_{HF}^2 \cdot K_2 \cdot (K_1 + K_1')}\right)\right), \quad \text{with } K_i = 10^{(\alpha \cdot \Delta G_i^0 + \beta)},\ K_1' = 10^{(\delta \cdot \Delta G_i^0 + \beta)} \text{ and } a_{HF} = \frac{1}{a \cdot (K_1 + K_1') \cdot K_2 + b} \tag{10}$$

As in the first case, two scenarios were proposed: one considering that all the formation constants are computed the same way [that is $\alpha = \delta$ in Eq. (10)] and the other removing this constraint to yield an extra degree of freedom, δ. The results for these models (4 and 5) are displayed in the third part of Table 4. The a_{HF} model had to be modified to take into account the effect on the acid concentration of the formation constants, K_1' and K_2' (the actual formula is $a_{HF} = 1/[a \cdot (K_1 \cdot K_2 + K_1' \cdot K_2') + b]$).

Of these two models, the latter was more accurate by some margin, presumably for the same reasons as discussed for case 1; the extra degree of freedom gives better fit, despite not being a rigorous experimental description. The improvement of the model 5 over Equation (6) and model 2 was impressive. Having extra degrees of freedom, the Fischer's *F*-statistic of this model was similar to the others. However, the overall correlation or cross-validated coefficients increase to excellent values and the root-mean-square error of this model is much closer to the experimental one (0.19 versus 0.10). An interesting point is the value of α in this model; in comparison to other models, the value decreased significantly, which could mean that other models get a higher value to compensate some underestimated predictions, especially for strong bases. Another interesting point is on the symmetry of the proposed hypothesis. In effect, instead of taking $K_2' = K_2$, we could take $K_2 = K_2'$ (meaning that K_2 and K_2' will be computed with δ instead of α). Surprisingly, the statistics are identical and the changes into the coefficients stay within the error limit (except for *b*, which doubles in value). This information supported our choice of hypothesis. Moreover, if we decide to take the more constraining hypothesis $\Delta G_2^0 = \Delta G_2^{0\prime}$, we get similar results ($n = 40$, $R^2 = 0.969$, $R_{CV}^2 = 0.966$, rms = 0.191, $F = 373.1$). The plot of observed versus calculated pK_{HB} values for model 5 [from Eq. (10)] is displayed in Figure 6.

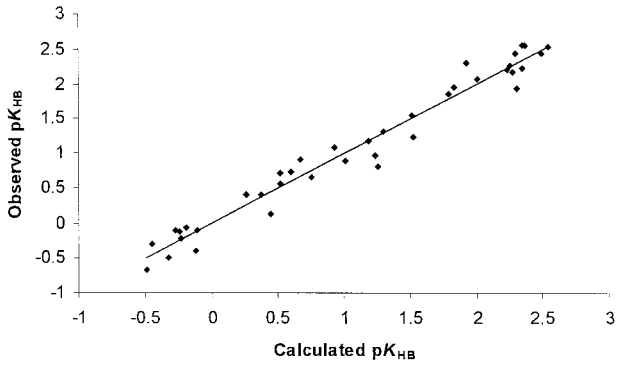

Figure 6. Observed versus calculated pK_{HB} from Equation (10), model 5 for the test set.

Linear fit using Gibbs free energy and indicator variable: From the models proposed above, we see that the most important corrections are made for the amide and sulfoxide families, strong carbonyl-based bases. It is tempting to propose a quick linear fit by applying an average correction only for these two families. A new indicator variable θ was therefore defined as unity for amide and sulfoxide compounds and zero otherwise. Using this new descriptor, we obtained a multivariate linear regression equation, which correlates pK_{HB} according to Equation (11).

$$pK_{HB} = 0.431(0.031) + 0.075(0.002) \cdot \Delta G^0 + 0.690(0.083) \cdot \theta \tag{11}$$

with $n = 40$, $R^2 = 0.975$, $R_{CV}^2 = 0.971$, rms = 0.166, $F = 728.7$

The Gibbs free energy used in this equation is for the most stable 1:1 complex [i.e., that used in Eq. (6)]. This Equation is similar to Equation (6), with an identical coefficient of ΔG^0 and an average pK_{HB} value (0.69) added to amide and sulfoxide compounds. These statistics are very impressive, but

perhaps should serve as a caution against drawing conclusions from fit statistics only. In effect, this model will be less flexible in terms of prediction for two reasons. Firstly, it is unclear whether to set θ to 0 or 1 for new classes of chemicals, for example phosphine oxides. Also, the average correction may not be appropriate in some cases (for example too much for a weak amide such dimethylcarbamoyl chloride, $pK_{HB} = 1.00$). To properly test the performance of each model, validation on an external set is necessary.

Validation of the proposed models—Prediction of pK_{HB} values from an external set

Validation from a large set molecules: Proper comparison of the models of pK_{HB} required one to test each model with an external set of molecules. A test set of 22 compounds (see Table 5) was selected from the experimental database of pK_{HB} values, with molecules selected to cover both the numerical spread and the chemical diversity of molecules of the proposed model. Some molecules were also selected to extend some existing family (ether and ketone) as well as the chemical diversity (e.g. aldehyde, phosphine oxide and nitro compounds). Other molecules were selected to test possible weaknesses of the model (e.g. 1-methyl-2-piperidone, dimethyl cyanamide or pyrimidine). Molecules with two equivalent acceptor sites had their apparent pK_{HB} value corrected by $-\log 2$ (e.g. biacetyl, pyrimidine or 1,3-difluoropropane). The value for nitromethane was also statistically corrected by $-\log 2$, as it is clear from the optimised structure that the hydrogen bond formed is two-centred. As discussed above, we found two stereoisomeric 1:1 complexes for the sp^2 and sp^3 hybrid oxygen sites except for methoxybenzene, which has, similarly to phenols, a planar 1:1 complex. The θ descriptor was set one for the amide, sulfoxide and phosphine oxide compounds and zero otherwise.

The predictive ability of the different models was tested using i) R^2 of correlation between predicted and observed pK_{HB} values; ii) the standard deviation of the residual pK_{HB} values, *sd*; iii) the average error, *ae*; iv) the average absolute error, *aae*; and v) the difference between the maximum and minimum residual pK_{HB} values, Max-Min. Statistics for each model are displayed in Table 6. The statistics on the predictions given by Equation (6) were significantly worse than the corresponding fit model. This was not surprising, as many compounds have been especially chosen to test the model on its weaknesses. The best predictive equation of Table 6 was model 5 (from Table 4), which seemed to be more robust than the others as its predictive statistics stay close to the ones of the fit model and better than the other models. Similarly, model 2 gave good statistics for predictions, and required much less time and computational resources. The last model of Table 6 had good statistics, similar to model 2, but considerably less than its corresponding fit [Eq. (11)] showing its lack of flexibility and robustness. Figure 7 is a plot of the observed pK_{HB} values versus calculated pK_{HB} values from model 5.

Table 5. Values of the hydrogen-bond basicity parameter for the validation test.

Number	Name	pK_{HB}
aldehyde[a]		
41	acetaldehyde	0.65
42	benzaldehyde	0.78
alkyl halide [b]		
43	1,3-difluoropropane	−0.27
carboxylic amide [c]		
44	dimethylcarbamoyl chloride	1.00
45	dimethyl formamide	2.10
46	1-methyl-2-pyrrolidone	2.38
47	1-methyl-2-piperidone	2.60
carboxylic ester[d]		
48	propiolactone	0.86
49	valerolactone	1.43
ether[e]		
50	methoxybenzene	0.02
51	2-chloroethylethyl ether	0.55
52	propylene oxide	0.97
53	diethyl ether	1.01
ketone[a]		
54	1,1,1-trifluoroacetone	−0.06
55	biacetyl	0.23
56	acetophenone	1.11
57	cyclohexanone	1.39
nitrile[f]		
58	cyanogen bromide	0.19
59	dimethyl cyanamide	1.56
pyridine[g]		
60	pyrimidine	1.07
phosphine oxide[h]		
61	trimethyl phosphate	2.50
nitro compound[i]		
62	nitromethane	−0.03

[a] F. Besseau, *J. Chem. Soc. Perkin Trans. 2* **1998**, 101. [b] C. Ouvrard, *J. Chem. Soc. Perkin Trans. 2* **1999**, 1357. [c] J. Y. Le Questel, *J. Chem. Soc. Perkin Trans. 2* **1992**, 2091. [d] F. Besseau, *J. Chem. Soc. Perkin Trans. 2* **1994**, 485. [e] M. Berthelot, *Eur. J. Org. Chem*. **1998**, 925. [f] a) J. Y. Le Questel, *J. Chem. Soc. Perkin Trans. 2* **1997**, 2711; b) M. Berthelot, *J. Chem. Soc. Perkin Trans. 2* **1993**, 625. [g] F. Besseau, *J. Chem. Soc. Perkin Trans. 2* **1998**, 283. [h] C. Laurence, unpublished results. [i] C. Laurence, *J. Chem. Soc. Perkin Trans. 2* **1994**, 491.

Table 6. Prediction statistics for the 22 validation molecules.

Model[a]	R^2	*sd*	*ae*	*aae*	Max-Min
linear fit from Eq. (6)	0.856	0.312	−0.115	0.244	1.13
model 2 from Table 4	0.934	0.212	−0.006	0.181	0.66
model 5 from Table 4	0.960	0.166	0.019	0.140	0.59
quick fit from Eq. (11)	0.937	0.206	−0.033	0.165	0.66

[a] See text for definition.

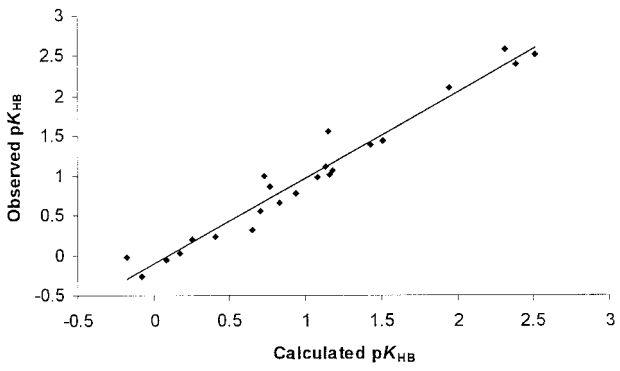

Figure 7. Observed versus calculated pK_{HB} from Equation (10), model 5 for the validation set.

0947-6539/02/0802-0464 $ 17.50+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 2

From the previous results and the ones displayed in Table 6, we conclude that the main strengths of the two best proposed models (models 2 and 5, from Table 4) were their ability to predict pK_{HB} values for a wide diversity of compounds and a large range of pK_{HB} values. Even new chemical structures not present in the training set (e.g. phosphine oxide or aldehyde) or molecules with cumulative difficulties (e.g. biacetyl) are predicted with similar accuracy. From the three models, Equation (6), models 2 and 5, we noticed that the quality of the model was proportional to the degree of complexity, but also the time required. A good compromise was therefore model 2, that taking into account only the second stereoisomeric 1:1 complex, as we obtained good predictions for a modest increase of the computational time. We also point out that both 1:1 complexes must be optimised to know which one would be required for Equation (6).

We have already discussed above (Section on 1:1 complexes) that only the complexes relative to the most stable acceptor atom were taken into account in the different models. It is notable that some compounds can have many other acceptor atoms (>2), such as trimethyl phosphate in our test set (three other sp^3 hybrid oxygen atoms). So, in order to corroborate the hypothesis proposed in the first Section, different calculations of pK_{HB} values have been done on trimethyl phosphate, taking into account the six other 1:1 complexes (stereoisomeric complexes). The pK_{HB} value obtained from Equation (6) was 1.63, and became 1.67 with the three extra formation constants, making the pK_{HB} difference appreciable (but still less than the rms error of this model). However, a better correction can be found by using a non-linear fit. Using model 2, the new pK_{HB} value obtained was 2.34 (using the Gibbs free energy of the two stereoisomeric complexes of the most stable hydrogen-bond site) and became 2.35 if we take into account all four hydrogen-bond sites (equivalent to eight formation constants in the logarithmic term). The pK_{HB} difference was now negligible, supporting the hypothesis made earlier. However, such cumulative effects could become important for molecules with many different secondary hydrogen-bond sites, such as polymers.

Validation from a set of molecules with intramolecular hydrogen bonding: Another interesting test for the proposed model was to predict the value of bases with different acceptor sites and intramolecular hydrogen bonding. Table 7 displays the different predicted pK_{HB} values, along with their experimental values. The equation used to calculate the pK_{HB} value is the one from model 2 (i.e., no 2:1 complexes considered). As discussed above, in the case of more than one equivalent hydrogen-bond accepting group, the measured formation constant has to be statistically corrected by -log2, as it is the sum of the different individual constants. This was the case with 1,4-diaminobutane, where the predicted value was in accordance with the statistically modified one. Smaller molecules have been found to be anomalously stronger bases than their analogues,[31] possibly as a result of intramolecular hydrogen bonding through five-membered and six-membered ring structures.[32] Calculations made on ethylenediamine and 2-methoxyethylamine show that the most stable conformation

Table 7. pK_{HB} predictions for bases with intramolecular hydrogen bonding.

Name	Calcd[a] pK_{HB}	Calcd[b] pK_{HB}	Obs. pK_{HB}	Obs.[c] pK_{HB}
1,4-diaminobutane[d]	**2.19**	–	2.51	**2.21**
ethylenediamine[d]	2.06	**2.51**	**2.55**	2.25
2-methoxyethylamine[d]	1.98	**2.26**	**2.29**	2.28

[a] No intramolecular hydrogen bond considered. [b] Intramolecular hydrogen bond considered. [c] Statistically corrected due to second hydrogen bond acceptor group. [d] J. Graton, *J. Chem. Soc. Perkin Trans. 2* **1999**, 997.

is indeed that with intramolecular hydrogen bond. The basicity of the acceptor site, as measured by ΔG^0 for the 1:1 complex, was enhanced by the intramolecular hydrogen bond, and excellent agreement with experiment found. No statistical correction is required here, as the second basic site is involved in intramolecular hydrogen bonding. It is encouraging to see that the model can cope easily with the introduction of intramolecular hydrogen bonding.

Construction of a model from training and test sets: Finally, we combined the training and test sets to give an overall training set of 62 molecules, and re-optimised the coefficients of the best models: The results are displayed in Table 8. It is reassuring that all regression coefficients were similar to those

Table 8. Nonlinear correlations with HF complex properties for 62 molecules.

Model[a]	Regression coefficients[b]					Statistics				
	α	β	δ	a	b	n	R^2	R_{CV}^2	rms	F-stat
Eq. (6)	0.084	0.498	–	–	–	62	0.913	0.907	0.286	626.9
	0.003	**0.045**	–	–	–					
model 2	0.078	0.356	0.123	–	–	62	0.950	0.947	0.218	560.8
	0.003	**0.033**	**0.007**	–	–					
model 5	0.054	0.195	0.097	0.032	10.97	62	0.966	0.964	0.181	552.8
	0.005	**0.042**	**0.011**	**0.007**	**3.89**					

[a] See text for definition. [b] The values in bold below the regression coefficients are the standard errors.

for 40 molecules, such that the models were robust to the introduction of new molecules. The fit statistics were almost equivalent to previous models, except for the linear fit as we introduced more compounds with 1:1 stereoisomeric complexes, which is the main weakness of the linear model. By extending the test set, we can conclude that the two best proposed models (models 2 and 5) are very robust and flexible as we introduced more challenging compounds and new chemical structures. The linear model is still reasonable, but its quality is likely to drop if we introduced more compounds with sp^2 and sp^3 hybrid oxygen atoms.

Conclusion

We have shown that density functional calculations using moderately sized basis sets can yield accurate models of the hydrogen-bond basicity, pK_{HB}. One property, the hydrogen-bond Gibbs free energy (ΔG^0), computed at the B3LYP/6-31+G(d,p) level using optimised geometries at this level, is

found to be significant in several models. ΔG^0 for the formation of the most stable 1:1 complex with hydrogen fluoride leads to a linear equation with reasonable overall statistics, but unacceptably large errors for some classes of compounds. Classes with stereoisomeric 1:1 complexes, such as amides and ethers, can have poor prediction.

The model can be improved by including the Gibbs free energies of hydrogen-bond formation for these stereoisomeric 1:1 complexes into a non-linear fit. This results in excellent statistics, with good overall correlation coefficient and a satisfactory root mean square error. Further improvements are possible by considering linear 2:1 complexes in addition to the aforementioned complexes, which yield an excellent overall correlation coefficient and a root-mean-square error close to experimental error. These three models increase in complexity, but also in time and computational resources required: we believe the best compromise between time and accuracy comes from the second model, considering the stereoisomeric 1:1 complexes. However, if accuracy is a key target, the model including as well the 2:1 complexes may be preferred. Finally, a quick multivariate linear model is proposed with an average correction for strong carbonyl-based bases: this is an accurate but less flexible and robust model than its predecessors.

Acknowledgements

O.L. thanks GlaxoSmithKline for a Ph.D. studentship. The authors are grateful to Prof. Christian Laurence (Nantes) for many helpful discussions.

[1] See for example: a) C. F. Mellot, A. M. Davidson, J. Eckert, A. K. Cheetham, *J. Phys. Chem. B* **1998**, *102*, 2530; b) S. E. Barber, K. E. S. Dean, A. J. Kirby, *Can. J. Chem.* **1999**, *77*, 792; c) C. Cox, T. Lectka, *J. Am. Chem. Soc.* **1998**, *120*, 10660.
[2] For recent reviews, see: a) A. Ivancich, T. A. Mattioli, *Photosynth. Res.* **1998**, *55*, 207; b) R. Aurora, G. D. Rose, *Protein Sci.* **1998**, *7*, 21.
[3] G. R. Desiraju, *Crystal Engineering—The Design of Organic Solids*, Elsevier, New York, **1989**.
[4] T. Steiner, G. R. Desiraju, *Chem. Commun.* **1998**, *8*, 891.
[5] a) D. E. Leahy, J. J. Morris, P. J. Taylor, A. R. Wait, *J. Chem. Soc. Perkin Trans. 2* **1992**, *4*, 705; b) M. Berthelot, C. Laurence, D. Foucher, R. W. Taft, *J. Phys. Org. Chem.* **1996**, *9*, 255.
[6] a) O. A. Raevsky, K. J. Schaper, *Eur. J. Med. Chem.* **1998**, *33*, 799; b) H. van de Waterbeemd, M. Kansy, *Chimia* **1992**, *46*, 299.
[7] a) S. Spange, A. Reuter, *Langmuir* **1999**, *15*, 141; b) M. F. Vitha, P. W. Carr, *J. Phys. Chem. B* **1998**, *102*, 1888; c) T. B. Lloyd, *Colloids Surfaces A* **1995**, *93*, 25.
[8] a) B. Welke, K. Ettlinger, M. Riederer, *Env. Sci. Technol.* **1998**, *32*, 1099; b) E. G. Sarraf, *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 2519.
[9] a) D. Gurka, R. W. Taft, *J. Am. Chem. Soc.* **1969**, *91*, 4794; b) R. W. Taft, D. Gurka, L. Joris, P. von R. Schleyer, J. W. Rashkys, *J. Am. Chem. Soc.* **1969**, *91*, 4801; c) L. Joris, J. Mitsky, R. W. Taft, *J. Am. Chem. Soc.* **1972**, *94*, 3438.
[10] See for example: J. Graton, C. Laurence, M. Berthelot, J. Y. Le Questel, F. Besseau, E. Raczynska, *J. Chem. Soc. Perkin Trans. 2* **1999**, *5*, 997.
[11] J. Y. Le Questel, M. Berthelot, C. Laurence, *J. Chem. Soc. Perkin Trans. 2* **1997**, *12*, 2711.
[12] J. S. Murray, P. Politzer, *J. Chem. Research (S)* **1992**, *3*, 110.
[13] H. Hagelin, J. S. Murray, T. Brinck, M. Berthelot, P. Politzer, *Can. J. Chem.* **1995**, *73*, 483.
[14] See for example: a) F. Besseau, C. Laurence, M. Berthelot, *J. Chem. Soc. Perkin Trans. 2* **1994**, *3*, 485; b) C. Laurence, M. Berthelot, M. Luçon, D. G. Morris, *J. Chem. Soc. Perkin Trans. 2* **1994**, *3*, 491.
[15] M. H. Abraham, P. L. Grellier, D. V. Prior, J. J. Morris, P. J. Taylor, *J. Chem. Soc. Perkin Trans. 2* **1990**, *4*, 521.
[16] a) M. H. Abraham, *Chem. Soc. Rev.* **1993**, *22*, 73; b) M. H. Abraham, *Pure Appl. Chem.* **1993**, *65*, 2503.
[17] J. A. Platts, *Phys. Chem. Chem. Phys.* **2000**, *2*, 3115.
[18] See for example: F. Besseau, M. Luçon, C. Laurence, M. Berthelot, *J. Chem. Soc. Perkin Trans. 2* **1998**, *1*, 101.
[19] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, J. A. Pople, *Gaussian 98*, Rev. A.6, Gaussian, Inc., Pittsburgh, PA (USA), **1998**.
[20] a) R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724; b) M. S. Gordon, *Chem. Phys. Lett.* **1980**, *76*, 163.
[21] a) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648; b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[22] M. J. S. Dewar, C. H. Reynolds, *J. Comput. Chem.* **1981**, *2*, 140.
[23] T. Clark, J. Chandrasekhar, G. W. Spitznagel, *J. Comput. Chem.* **1983**, *4*, 294.
[24] S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117.
[25] JMP, Revision 4, SAS Institute, Inc., Cary, NC (USA), **2000**.
[26] a) C. Laurence, unpublished results; b) see for example: M. Berthelot, F. Besseau, C. Laurence, *Eur. J. Org. Chem.* **1998**, 925.
[27] a) H. Fritzsche, *Spectrochim. Acta* **1965**, *21*, 799; b) J. Korppi-tommola, H. F. Shurvell, *Can. J. Chem.* **1978**, *56*, 2959; c) J. Korppi-tommola, H. F. Shurvell, *Can. J. Chem.* **1979**, *57*, 2707.
[28] C. Laurence, M. Berthelot, M. Helbert, *Spectrochim. Acta Part A* **1985**, *41*, 883.
[29] A. Massat, P. Guillaume, J. P. Doucet, J. E. Dubois, *J. Mol. Struct.* **1991**, *244*, 69.
[30] M. H. Abraham, D. V. Prior, R. A. Schulz, J. J. Morris, P. J. Taylor, *J. Chem. Soc. Faraday Trans.* **1998**, *94*, 879.
[31] P. L. Huyskens, *J. Am. Chem. Soc.* **1977**, *99*, 2578.
[32] a) K. M. Marstokk, H. Mllendal, *J. Mol. Struct.* **1978**, *49*, 221; b) W. Caminati, E. B. Wilson, *J. Mol. Spectrosc.* **1980**, *81*, 356.

Received: July 2, 2001 [F3384]

